LAST NAME (Please Print): KEY

FIRST NAME (Please Print): _____

HONOR PLEDGE (Please Sign): _____

Statistics 111

Midterm 1

- This is a closed book exam.
- You may use your calculator and a single page of notes.
- The room is crowded. Please be careful to look only at your own exam. Try to sit one seat apart; the proctors may ask you to randomize your seating a bit.
- Report all numerical answers to at least two correct decimal places or (when appropriate) write them as a fraction.
- All question parts count for 1 point.

1. Springfield Elementary has 500 students; 200 are girls and 300 are boys. For a school trip, the children could sign up to visit either the circus or the zoo. 150 girls and 50 boys went to the zoo; the rest saw the circus. Predictably, some students behaved badly on the field trip. Specifically, 40% of the children at the zoo misbehaved, of whom 5 were boys; but only 25% of the children at the circus misbehaved, and of them 50 were boys. Principal Skinner is planning to cancel the trip to the zoo next year, on the grounds that its promotes mischief, but Lisa opposes him.

It is helpful to write the information above as a table, shown below:

	Boys			Girls		
	Behave	Misbehave		Behave	Misbehave	
Zoo	45	5	50	75	75	150
Circus	200	50	250	25	25	50
	300			200		

In the context of the Berkeley admissions problem, acceptance or rejection corresponds to behaving or misbehaving, gender corresponds to the zoo or circus, and major corresponds to boy or girl.

40% What is the misbehavior rate at the zoo? (5+75)/(50+150).

10% What is the misbehavior rate for boys at the zoo? 5/50

20% What is the misbehavior rate for boys at the circus? 50/250

Assume that girls behave equally well at the zoo and circus. What is Lisa's explanation to Principal Skinner?

It is Simpson's paradox. Girls tend to behave much more badly than boys, and girls strongly prefer to go to the zoo. So it is not the zoo that is promoting uncouth behavior, but gender.

32% What is the corrected misbehavior rate at the circus?

The formula is proportion of boys * misbehavior rate of boys at the circus + proportion of girls * misbehavior rate of girls at the circus = (300/500) * (0.2) + (200/500) * (0.5) = 0.32.

0.44 2. In a class of 105 students, what is the approximate probability that 14 or fewer were born on Monday? (Assume all days are equally likely.)

Normal approximation to the Binomial. The mean is 15 and the sd $\sqrt{105 * (1/7) * (6/7)} = 3.5857$. So with the continuity correction, $z = (14.5 - 15)/3.5857 = -0.1394$. From the table, the probability is 0.4443.

0.05 3. On average, there are 0.8 bank robberies in North Carolina per day. What is the probability that there are more than 2 robberies tomorrow?

Poisson. $1 - \exp(-0.8)\left(\frac{0.8^0}{0!} + \frac{0.8^1}{1!} + \frac{0.8^2}{2!}\right) = 0.0474$.

0.25 4. Consider the density function $f(x) = c|x|$ on the interval $[-2, 2]$, 0 else. What is c?

The area under the density is 1. The triangle on the right has 2, so does the one on the left, so $c = 1/4$.

0.33 5. Let $F(x) = x^4/16$ on $[0, 2]$. What is the standard deviation of X?

$f(x) = x^3/4$ on $[0, 2]$ so $\mathbb{E}[X] = \int_0^2 x * x^3/4\, dx = 1.6$. And $\mathbb{E}[X^2] = \int_0^2 x^2 * x^3/4\, dx = 2.6667$. So the standard deviation is $\sqrt{2.667 - (1.6)^2} = 0.3266$.

6. Consider the following numbers (a sample from a population):

$$10,\ \ 0,\ \ 9,\ \ -10,\ \ 4,\ \ 8,\ \ 4,\ \ 2$$

7.5 What is the IQR?

4 What is the mode?

41.41 What is the sample variance? Divide by $n - 1$.

8 Suppose a sample has IQR equal to 4. You add 4 to each number and multiply by -2. What is the new IQR?

-2 Suppose a set of numbers has mean equal to 4. You subtract 3 from each number and multiply by -2. What is the new mean?

7. A box of chocalate contains dark, milk and white chocolate candies. For each type of chocolate, there is one that contains a cherry, one that contains a peanut, one that contains coconut, and one that contains nougat.

0.5 You eat one at random. What is the probability that it is dark chocolate or contains a cherry?

4/12 + 3/12 - 1/12 = 0.5

0.25 You eat three at random. What is the probability that none are dark chocolate?

8/12 * 7/11 * 6/10 = 0.2545.

0.32 You eat two at random. What is the probability that the second is white chocolate if the first was not dark chocolate?

The denominator is 2/3, the probability that the first is not dark chocolate. The numerator is P[(w,w)] + P[(m,w)] = (4/12) * (3/11) + (4/12) * (4/11) = 0.2121, so the ratio is 0.3182.

0.66 8. Suppose 20% of Duke students major in Econ, 30% of students are from NC, and 40% are male. If you lose your ID and a Duke student returns it, what is the probability that the person is an Econ major or from NC or male? (Hint: Venn diagram.)

0.2 + 0.3 + 0.4 - (0.2 * 0.3) - (0.2 * 0.4) - (0.3 * 0.4) + (0.2 * 0.3 * 0.4) = 0.664.

0.68 9. A university has 15 residence halls, all of equal size. If 6 students are drawn at random, what is the probability that two or more are in the same residence hall?

1 - P[none are from same hall] = 1 - 15/15 * 14/15 * 13/15 * 12/15 * 11/15 * 10/15 = 0.6836.

10. Let $f(x,y) = 3x/8$ for $0 \le y \le x \le 2$.

What is the marginal density of X? Indicate the support.

$$f_1(x) = \int_0^x 3x/8\, dy = 3x^2/8 \text{ for } 0 \le x \le 2, \text{ and } 0else.$$

What is the conditional density of Y given $X = x$? Indicate the support.

$$g_2(y|x) = f(x,y)/f_1(x) = 1/x \text{ for } 0 \le y \le x, \text{ and } 0else.$$

0.67 What is the probability that $Y \le 1$ given that X is 1.5?

$$\mathbb{P}[Y \le 1 | X = 1.5] = \int_0^1 g_2(y|x = 1.5)\, dy = 2/3.$$

11. You toss a coin that has probability 1/4 of coming up Heads. If it comes up Heads you draw a random value from a Poisson distribution with $\lambda_1 = 1.5$. If it comes up tails you draw a random value from an exponential with $\lambda_2 = 3.5$.

0.59 What is the expected value of your random variable?

$\mathbb{E}[X] = (1/4) * 1.5 + (3/4) * (1/3.5) = 0.5892.$

0.40 If you observe $X > \frac{1}{2}$, what is the probability that you threw Tails?

Bayes rule. $\mathbb{P}[T|X > 0.5] = \frac{\mathbb{P}[X>0.5|T]*(3/4)}{\mathbb{P}[X>0.5|T]*(3/4)+\mathbb{P}[X>0.5|H]*(1/4)}$ where $\mathbb{P}[X > 0.5|T] = \exp(-3.5*0.5)$ and $\mathbb{P}[X > 0.5|H] = 1-\exp(-1.5)$. Putting this together gives 0.4016.

_____ 12. Suppose 20% of Duke students major in Econ. You go on six random dates. Let X be the number of Econ majors in the first four dates, and let Y be the total number of Econ majors among all the dates. What is Cov(X, Y)?

By definition, $\text{Cov}(X,Y) = \mathbb{E}[X*Y] - \mathbb{E}[X]*\mathbb{E}[Y]$. The mean of a binomial is np, so $\mathbb{E}[X] = 4p$ and $\mathbb{E}[Y] = 6p$. Now write $Y = X + Z$ where Z is the number of Econ majors among the last two dates. Then Z is independent of X and $\mathbb{E}[XY] = \mathbb{E}[X*(X+Z)] = \mathbb{E}[X^2 + XZ] = \mathbb{E}[X^2] + \mathbb{E}[XZ] = \mathbb{E}[X^2] + \mathbb{E}[X]\mathbb{E}[Z]$. We know that $\mathbb{E}[Z] = 2p$. And since the variance of a binomial is $np(1-p) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$, then $4p(1-p) = \mathbb{E}[X^2] - (4p)^2$ and solving gives $\mathbb{E}[X^2] = 4p + 12p^2$. So the covariance is $4p - 4p^2$. For $p = 0.2$ this is

0.25 13. Suppose the height of a Frisian male is normally distributed with mean 78 inches and an sd of 3 inches. (Dante thought they were the tallest people in the world.) What is the probability that Henk, a Frisian, is more than 80 inches tall?

$z = (80 - 78)/3 = 0.6667$. From the table, the answer is 0.2514.

0.61 14. At least what proportion of the data from any distribution must be within 1.6 standard deviations of the mean?

Tchebyshev. $1 - (1/1.6)^2 = 0.6093$.

15. Suppose $f(x, y) = c$ for (x, y) inside the diamond bounded by (-1, 3), (1, 5), (3, 3) and (1, 0), and it is zero elsewhere.

1/10 What is c?

The upper triangle has base 4 and height 2 for area 4. The lower triangle hs base 4 and height 3, for area 6. So the total area is 10.

No Are X and Y independent?

If $x = 3$ then you know that y must be 3.

What is $f_1(x)$? Indicate the support.

Find $\int(1/10)dy$. For the triangle formed by (1,5), (1, 0) and (-1, 3), the upper limit of y is $x+4$ and the lower limit of y is $-3x/2+3/2$. So on that region $f_1(x) = (x+1)/4$.
For the triangle formed by (1, 5), (3, 3) and (1, 0), the upper limit of y is $6-x$ and the lower limit is $3x/2-3/2$. So on that region, $f_1(x) = (3-x)/4$. Thus:

$$f_1(x) = \begin{cases} (x+1)/4 & \text{if } -1 \le x \le 1 \\ (3-x)/4 & \text{if } 1 < x \le 3 \\ 0 & \text{else} \end{cases}$$

1 What is $\mathbb{E}[X]$?

By the symmetry of the diamond.

What is $g_1(x|y)$? Indicate the support.

The first step is to find $f_2(y) = \int(1/10)\,dx$. On the triangle formed by (-1,3), (1, 5) and (3, 3), the largest x can be is $6-y$ and the smallest x can be is $y-4$. So for $3 < y \le 5$, $f_2(y) = 1-(2y)/5$. And on the triangle formed by (-1, 3), (3, 3), and (1, 0), the largest value x can take is $(2y)/3-1$ and the smallest is $1-(2y)/3$. So for $0 \le y \le 3$, $f_2(y) = 2y/15$.
The second step is to find $g_1(x|y) = f(x,y)/f_2(y)$. The numerator is just 1/10, so

$$g_1(x|y) = \begin{cases} 3/(4y) & \text{if } 1-(2/3)y \le x \le 1+(2/3)y \\ 1/[2(5-2y)] & \text{if } y-4 < x \le 6-y \\ 0 & \text{else} \end{cases}$$

0 What is the correlation between X and Y?

We can add a constant to X and Y without changing the covariance. So recenter the diamond at (0, 0). Then, by symmetry, $\mathbb{E}[XY] = 0$. And although $\mathbb{E}[Y] \neq 0$, symmetry implies $\mathbb{E}[X] = 0$. So $\text{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$, forcing the correlation to be 0.

-4.8 16. Suppose X has the normal distribution with mean 6 and standard deviation 3, and that Y has the normal distribution with mean -3 and variance 4. Suppose their correlation is -0.8. What is their covariance?

$\text{Corr}(X,Y) = \text{Cov}(X,Y)/(sd_x * sd_y)$, so the covariance is -4.8.

median 17. Suppose Duke required the sumo wrestling team to take STA 111 in the spring. You measure all the weights in the class. Which will be lower—the mean or the median?

The median, since the mean is pulled up by outliers.

0.44 18. A class consists entirely of econ majors and statistics majors (and no double majors). There are 6 econ majors and 7 statistics majors. You choose four students at random without replacement. What is the probability that you have exactly two of each?

Hypergeometric. 0.4406.